

Anomalies detection in real data from experimental flood embankment

Monika Chuchro, Kamil Szostek, Andrzej Leśniak
{chuchro, szostek, lesniak}@agh.edu.pl

AGH University of Science and Technology, Department of Geoinformatics and Applied Computer Science

Introduction

The aim of ISMOP project is to conduct comprehensive research on the system for monitoring and forecasting the static and dynamic state of flood embankment. The project has developed methods of massive data collection from various sensors in continuous mode, effective saving into database, interpretation and analysis of data (including the use of numerical modelling). One of the most important goals of the project is to provide information about the state of the flood embankment with visualization (fig.1) of the experimental flood embankment located in Czernichow [1,2,3].

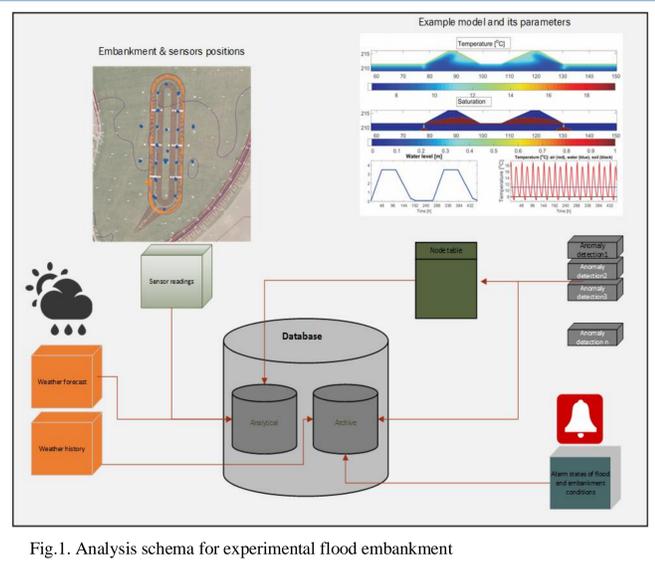


Fig.1. Analysis schema for experimental flood embankment

Data-driven module

The method for assessing an embankment stability is the analysis of data from sensors, aiming to detect changes and deviations from standard values. Temperature and pore pressure time series are periodic, with very weak daily periodicity and explicit periodicity associated with the seasons. A characteristic feature of the analysed time series is the lack of a trend and the strong influence of irregular components related to weather conditions.

The aim of the analysis is to detect emerging anomalies at the end of the analysed time. Anomalies might be present as group values higher than the average values, a huge single value, or a growing trend. Such changes in the average level of the phenomenon, referred to as anomalies, are an indicator of unfavourable changes in the flood embankment that could indicate instability.

Anomaly detection in a time series of sensor measurements was achieved using methods based on Fast Fourier Transform (FFT) and frequency models. In anomaly detection, two time series of the same length and the same sampling step are analysed and recorded from one sensor in similar atmospheric conditions. One of the time series is a data set for the absence or presence of anomalies is known. The second time series is a data set that we want to test for the occurrence of anomalies. In order to detect anomalies, spectral density values are compared for the first model. In the second model, for which a frequency model is calculated, changes in adjustment are rated, expressed by coefficient of determination (R^2), distribution of residuals, and changes in residuals variance (fig.2). If the difference between the assessed parameters exceeds the critical value in the analysed time series, an anomaly is detected. The second case is when comparable parameters are not significantly different from each other and an anomaly was detected in a series that was compared, then the analysed time series probably contains an anomaly [4,5].

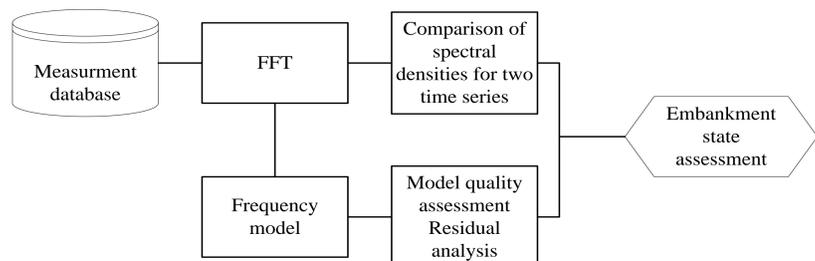


Fig.2. General scheme of the data-driven module

Bibliography

1. M. Bubak, B. Baliś, D. Hareźlak, M. Kasztelnik, P. Nowakowski, T. Bartyński, T. Gubała, M. Malawski, M. Pawlik, B. Wilk: The ISMOP IT platform for smart levee monitoring and flood decision support, KU KDM 2016 : ninth ACC Cyfronet AGH HPC users' conference : Zakopane 16–18 March 2016 : proceedings, pp. 35–36, 2016.
2. K. Wiatr, J. Kitowski, M. Bubak (Eds): An approach to monitoring, data analytics, and decision support for levee supervision" Proceedings of the Seventh ACC Cyfronet AGH Users' Conference, ACC CYFRONET AGH, Kraków, ISBN 978-83-61433-09-5, pp. 75-76, 2014.
3. J. Stanisław, A. Borecka, A. Leśniak, K. Zieliński: Wybrane systemy monitorujące obwałowania przeciwpowodziowe, Przegląd Geologiczny, vol. 62, nr 10/2, pp. 699-703, 2014.
4. M. Chuchro, M. Lupa, B. Bukowska-Belniak, A. Leśniak: Detekcja potencjalnych anomalii pomiarów parametrów w wale przeciwpowodziowym., Studia Informatica, 37(1), pp. 175-185, 2016.
5. A. Piórkowski, A. Leśniak: Using data stream management systems in the design of monitoring system for flood embankments. Studia Informatica, 35(2), pp. 297-310, 2014.
6. M. Chuchro, M. Lupa, A. Pięta, A. Piórkowski, A. Leśniak: A concept of time windows length selection in stream databases in the context of sensor networks monitoring, New trends in database and information systems II, pp. 173-183, 2014.

Acknowledgments. This work was partially supported by the National Centre for Research and Development (NCBiR) under Grant No. PBS1/B9/18/2013 and by the AGH grant no. 11.11.140.630.

Applications

The application was written using C/C++ and compiled using Microsoft Visual C++ (MSVC) version 120 compiler and GCC 4.8 for Linux OS. The application is a batch program, which means it is evaluated using a command line. Therefore, it may be easily parallelized using multiple physical or virtual machines and used in more complex system for embankment monitoring.

The application uses data from two CSV input files. These files contain values of real measurements of temperatures and pore pressure from the embankment. The occurrence of anomaly is known only for the first one. The second file is then analyzed. An example of the application call with parameters and its results is presented below. The parameters, besides the input files paths, the FFT model frequencies were supplied. The output depicts that the second approach might have detected anomaly, because of the noticeably change in the SW-W coefficient [4,6].

```
$>anomalies.exe -a 3 -if data\Tout12_bez_anomali01062015h1300.csv -cf data\Tout12_01072015h1300.csv -vsT 0.00520833333,0.0104166667,0.015625,0.0208333333,0.0260416667,0.03125,0.0364583333,0.0416666667
```

```
Anomaly detection - method 3  
Coefficients of determination:  
Temp 0x2 (9.080025e-001 -> 9.005312e-001)  
SW-W:  
Temp 0x1 (0.987137 -> 0.693760)
```

Test

The anomalies detection algorithm method 1 was tested for measurements of pore pressure sensors UT6 to UT10 of the half section NW of experimental flood embankment, for the period from 2015-08-10 to 2016-09-13. Time period from 2015-08-10 to 2015-08-12 (192 observations) was selected as a period without anomaly. During this time the experimental flood embankment was dry. Since August 2016 have been performed flooding experiments. Algorithms were tested on pore pressure time series retested by sensors for period without water in flood embankment and with water simulated flood wave. For period without water in flood embankment tests should prove the absence of anomalies. For the second part of data (August and September 2016) tests should prove the occurrence of anomalies.

The 47734 iterations of the program were performed for each sensor. Anomalies in sensor UT6 were detected for 456 time periods. In the subsequent sensors, localized in a greater distance from water, number of anomalies decrease to 200 for sensor UT10. A sample period of time, for which the anomalies were detected, is shown below in Figure 3 and Figure 4. Data-driven method 3 also confirmed anomalies occurrence in periods pointed by method 1.

Data-driven method 2, analysis the shift in phase for two sensors from one half cross section. For chosen threshold values, the method 2 did not detect anomalies in periods pointed by method 1 and 3 (fig.3, fig.4).

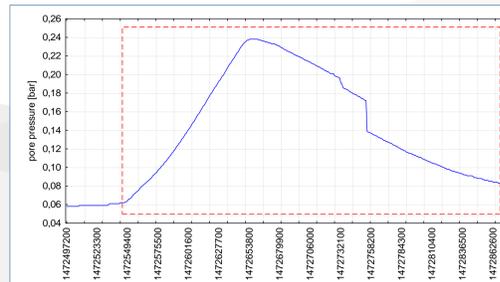


Fig.3. Sample detected anomalies, sensor UT6

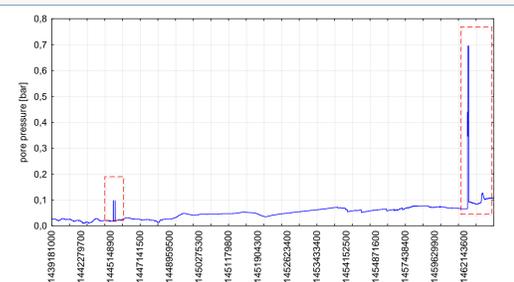


Fig.4. Sample detected anomalies, sensor UT7

Conclusions

The selected algorithms detected 456 potential anomalies for the sensor UT6 for period with flooding experiments (fig.3). For the UT7 sensor, algorithms detected two periods with anomaly, in the first part of time series (period without experiments) and for the period with flooding experiment, the same period as for UT6 sensor (fig.4).

The problem to solve is the proper selection of the algorithms parameters and critical values for distinguishing normal and abnormal values (or their changes), indicating the presence of potential anomalies.



The academic-industrial consortium:



www.ismop.edu.pl